
Proceedings Paper:
Zhu, Haoran, Wu, Hui, Wang, Jianjia et al. (1 more author) (2021) Weighted Network Analysis Using the Debye Model. In: Torsello, Andrea, Rossi, Luca, Pelillo, Marcello, Biggio, Battista and Robles-Kelly, Antonio, (eds.) Structural, Syntactic, and Statistical Pattern Recognition. Springer , Cham , pp. 153-163.

https://doi.org/10.1007/978-1-4419-8819-5

Weighted Network Analysis using the Debye Model

Haoran Zhu[1], Hui Wu[1] Jianjia Wang[1]*, and Edwin R. Hancock[2]

School of Computer Engineering and Science,
Shanghai University, Shanghai, China
Department of Computer Science, University of York, York, UK
*Corresponding Author: jianjiawang@shu.edu.cn

Abstract. Statistical mechanics provides effective means for complex network analysis, and in particular the classical Boltzmann partition function has been extensively used to explore network structure. One of the shortcomings of this model is that it is couched in terms of unweighted edges. To overcome this problem and to extend the utility of this type of analysis, in this paper, we explore how the Debye solid model can be used to describe the probability density function for particles in such a system. According to our analogy the distribution of node degree and edge-weight in the network can be derived from the distribution of molecular energy in the Debye model. This allows us to derive a probability density function for nodes, and thus is identical to the degree distribution for the case of uniformly weighted edges. We also consider the case where the edge weights follow a distribution (non-uniformly weighted edges). The corresponding network energy is the cumulative distribution function for the node degree. This distribution reveals a phase transition for the temperature dependence. The Debye model thus provides a new way to describe the node degree distribution in both unweighted and weighted networks.

Keywords: Debye's solid model, Degree Distribution, Weighted Networks

1 Introduction

The study of complex networks has attracted sustained interest since it allows the otherwise intractable interactions between the different units of complex systems to be represented and analysed [8]. This usually involves the study of the unweighted or weighted "edges" between vertices using methods from graph theory [4]. However, the node degree distribution also plays a critical role, since it describes the topological structure of networks and may determine the evolution characteristics of a network [2]. It is widely confirmed that many different types of real-world network exhibit a power-law degree distribution and this can be induced by a linear preferential evolution mechanism [1]. This property illuminates the statistical nature of structural connections in a network.

However, the literature mainly focusses on the analysis of the degree distribution for unweighted networks, and rarely considers the distribution of edge weights. This limits the exploration of the nature of network structure based on information concerning the distribution of edge degree combinations or edge-weights. Recently, sophisticated tools from statistical physics have provided powerful ways to extend this kind of analysis [6, 5]. These computationally efficient methods rely on thermodynamic analogies to describe the different structural or topological properties of networks [3]. For example, the Boltzmann distribution provides expressions for the macroscopic thermal characteristics such as temperature, energy and entropy from a microcosmic point of view [7]. This provides a novel framework to analyse and understand the statistical structural properties in weighted networks.

This aims to establish effective statistical mechanical methods for measuring the probability density function for nodes (and node degree) in weighted networks. We commence from a thermal analogy using Boltzmann statistics, which provides a physical meaning of the temperature and energy states in a network. This allows us to introduce and leverage the Debye solid model to calculate the degree distribution.

The Debye solid model originates is a statistical mechanical tool for the analysis of the distribution of phonon energy lattice structures from solid state physics. Specifically, it considers the vibrations (or phonons) of the atomic lattice. This treats the solid as an ensemble of harmonic oscillators. The model exhibits similar connectivity patterns to those found in complex networks. The connected nodes are analogous to the atoms, and the edge weights can be regarded as the phonon energies of the harmonic oscillators. Since, in this more general thermal analogy, the degree in the network has two degrees of freedom, i.e., in-degree and out-degree, the model builds on analogies with two a dimensional crystal.

Using this model, we find that for a given distribution of edge weights, the node probability in a weighted network not only depends on the node degree but also on the global temperature parameter. Furthermore, the corresponding network energy is just the cumulative distribution function for the node probability. Moreover, this reveals a phase transition for the temperature dependence.

2 Graph Representation

2.1 Preliminaries

Let $G(V, E)$ be an undirected graph with node set V and edge set $E \subseteq V \times V$. The edge-set can be represented by an adjacency A, with elements $A(u, v) = 1$ if $(u, v) \in E$ and $A(u, v) = 0$ otherwise. The diagonal degree matrix D has diagonal elements $D(u, u) = d(u)$, where $d_u = \sum_{v \in V} A_{uv}$ is the degree of node u, and off diagonal elements $D(u, v) = 0$ if $u \neq v$. Then, the Laplacian matrix is given by $L = D - A$.

For a weighted network G_w, the pair of nodes (u, v) has an associated real non-negative weight $w(u, v)$ for each edge, i.e., $u \in V, v \in V$, and $u \neq v$. The

adjacency matrix A_w for a weighted network is given by

$$A_w = \begin{cases} w(u,v) & \text{if } (u,v) \in E \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

where, for the undirected network, the weighted adjacency is symmetric, i.e., $w(u,v) = w(v,u)$ for all pairs of nodes such that $(u,v) \in E, u \neq v$.

2.2 Thermodynamic Representation

Here to model networks using a thermal analogy based on Boltzmann statistics, each network is regarded as an isolated system with a fixed number of both nodes $|V|$ and edges $|E|$. The nodes in the network are mapped onto the particles in the thermal system. Each edge has a unit weight. The corresponding node degrees are analogous to the discrete energy states. and the energy associated with each node is proportional to the node degree, that is

$$\omega_u = \varepsilon k \tag{2}$$

where ω_u is the energy per node which is identical to the node weight; and $\varepsilon = 1$ for an unweighted network, k is the degree per node; and $k \in \mathbf{Z}$ which is a positive integer or zero and equal to the number of edges connecting to the node u. Thus, the occupation number of the energy states depends on the degree of the nodes connected by edges.

According to the Boltzmann distribution, the probability for an individual node to be at a particular energy state is given by the exponential function

$$P_u = \frac{1}{Z} e^{-\beta \omega_u} \tag{3}$$

where Z is the partition function subject to the constraint of energy conservation and given by

$$Z = \sum_{u=0}^{|V|} e^{-\beta \omega_u} \tag{4}$$

The average energy then can be derived from the Boltzmann partition function

$$\bar{U} = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} = -\frac{\partial \log Z}{\partial \beta} \tag{5}$$

This allows us to treat a network as a statistical ensemble with associated thermal properties such as a partition function and a total energy.

3 Statistical Ensembles

For a network subject to Boltzmann statistics and in thermal equilibrium with a fixed number of nodes and edges, the entropy can be computed using Boltzmann's law, i.e $S = \kappa_B \log W(U)$, where $W(U)$ is the multiplicity of states and the total energy in the network is

$$U = \varepsilon |E| \tag{6}$$

which is an integer number equal to the total number of edges when the weight is unity.

The entropy relates to the number of ways for choosing $|E|$ edges among the available $U + |V| - 1$ possibilities. This is given by the combinatorial formula in terms of the factorials

$$W(U) = \frac{(U + |V| - 1)!}{U!(|V| - 1)!} \tag{7}$$

When number of nodes and edges are large, then the expression $\log W(U)$ can be simplified by using Stirling's approximation $\log n! \approx n \log n$ and as a result

$$\begin{aligned} S &= \kappa_B \ln W \\ &= \log[(U + |V| - 1)!] - \log(U!) - \log[(|V| - 1)!] \\ &= (U + |V| - 1)\log(U + |V| - 1) - U \log U - (|V| - 1)\log(|V| - 1) \end{aligned} \tag{8}$$

where κ_B is the Boltzmann constant.

For a thermodynamic system of constant volume, the temperature (or equivalently the parameter β, i.e., the inverse temperature) is the rate of change of energy with respect to entropy of the network. That is given by

$$\beta = \left(\frac{\partial S}{\partial U}\right)_{|V|} = \frac{1}{w} \log \frac{U + |V| - 1}{U} \tag{9}$$

Given the temperature the partition function for the equilibrium state of the thermal network system can be represented by the series expansion

$$Z = \sum_{u=0}^{|V|} e^{-\beta \omega_u} = \frac{1 - e^{-|V|\beta\omega}}{1 - e^{-\beta\omega}} \approx \frac{1}{1 - e^{-\beta\omega}} \tag{10}$$

From Eq.(3), the probability for a given node at a particular energy state depends on the node degree

$$P(d_u = k) = \frac{1}{Z} e^{-\beta\omega_u} = \left(1 - e^{-\beta\omega}\right) e^{-\beta\varepsilon k} \tag{11}$$

This leads to definitions of energy and entropy that associated with the network structure.

4 The Debye Model

The above analysis makes the rather limiting assumption that the weight for each edge is uniform and the energy states for each node are discrete. It effectively assumes that the density of states is simply a delta function. It is better to assume a distribution of edge weights to make the nodal energy continuous by replacing a density distribution.

4.1 Node Probability

Hence, we would like to incorporate a function $g(\omega)$ which describes the density of edge weights to allow us to make a more detailed analysis. The number of edge states with weights between ω and $\omega + d\omega$ is given by $g(\omega)d\omega$ and we require that the total edge weights sums to to the number of edges, i.e. is given by

$$\int g(\omega) d\omega = |E| \tag{12}$$

Eq.(2) is equivalent to assuming that the node energy corresponds to the degree. Here,on the other hand, we allow a more complex vectorial representation which accommodates the more general case of directed networks, which admits both node in-degree and out-degree.

For the space of node in-degree and out-degree, we require two integers to specify each node, i.e. the probability density for each node is bivariate depending on two variables k_{in} and k_{out} and is normalised by the sum order these two integers or equivalently by an integral over the volume element dk_{in}, dk_{out} in the node. The discrete summation can thus be rewritten as the integral, that is

$$\sum_{k}(\cdots) = \frac{1}{4}\int_{0}^{\infty} 2\pi k dk(\cdots) \tag{13}$$

Then, the density of states per node as a function of k is given by

$$g(k)dk = \frac{S}{(2\pi)^2}\cdot 2\pi k dk \cdot 2 = \frac{Sk}{\pi}dk \tag{14}$$

where the nodes in a network are assumed to be the square of area $S = |V|^2$ and the factor 2 corresponds to the two degrees of freedom for edges.

Thus, the corresponding density of weights for each node is given by

$$g(\omega)d\omega = \frac{S}{\pi\varepsilon^2}\omega d\omega \tag{15}$$

To derive the thermal quantities in the Debye model as a function of temperature, we begin by writing down the logarithm of partition function as follows,

$$\log Z = \int_{0}^{\omega_T} g(\omega)d\omega \log\left[\frac{1}{1-e^{-\beta w}}\right] = -\int_{0}^{\omega_T} g(\omega)d\omega \log\left[1-e^{-\beta w}\right] \tag{16}$$

Then, from Eq.(5), we can calculate the energy of the network using

$$U = -\frac{\partial \log Z}{\partial \beta} = \int_{0}^{\omega_T} g(\omega)d\omega \cdot \frac{\omega}{e^{\beta\omega}-1} = \frac{S}{\pi\varepsilon^2}\int_{0}^{\omega_T}\frac{\omega^2}{e^{\beta\omega}-1}d\omega \tag{17}$$

Substituting Eq.(2) into Eq.(17), the corresponding energy is related to the degree and is given by

$$U = \int_{0}^{\omega_T}\frac{S\varepsilon}{\pi}\cdot\frac{k^2}{e^{\beta\varepsilon k}-1}dk = \int_{0}^{\omega_T} P(\beta,k)dk \tag{18}$$

As a result the probability of each node given the degree k and temperature β is

$$P(\beta,k) = \frac{S\varepsilon}{\pi}\cdot\frac{k^2}{e^{\beta\varepsilon k}-1} = \frac{|V|^2}{2\pi|E|}\cdot\frac{k^2}{e^{\beta\varepsilon k}-1} \tag{19}$$

where $S = |V|^2, U = 2|E|\varepsilon$. This describes the degree distribution in the weighted network. It not only relates to the node degree, but also depends on the global temperature parameter as well.

4.2 Upper Weight Boundary

Because there is an limit on the total number of edges given the number of nodes in the network, the weight distribution has an upper bound ω_T. This is defined by

$$\int_0^{\omega_T} g(\omega)d\omega = 2|E| \tag{20}$$

which, using Eq.(15), implies that

$$\omega_T = \left(4\pi\frac{|E|}{|V|^2}\right)^{1/2}\varepsilon \tag{21}$$

This allows us to rewrite Eq.(15) as

$$g(\omega)d\omega = \frac{4|E|\omega}{\omega_T^2}d\omega \tag{22}$$

Thus, we now have all the ingredients necessary to apply the Debye model to derive the macroscopic thermal characterisations for the network.

4.3 High- and Low-temperature Limits

Now we substitute Eq.(22) into Eq.(16) to write the logarithm of the partition function as

$$\log Z = -\frac{4|E|}{\omega_T^2}\int_0^{\omega_T}\omega\log\left[1-e^{-\beta w}\right]d\omega \tag{23}$$

According to Eq.(7), the average energy is

$$\bar{U} = \frac{4|E|}{\omega_T^2}\int_0^{\omega_T}\frac{\omega^2}{e^{\beta w}-1}d\omega = \frac{4|E|}{\omega_T^2\beta^3}\int_0^{\frac{x_T}{\beta}}\frac{x^2}{e^x-1}dx \tag{24}$$

where $x = \beta\omega = \beta\varepsilon k$. This equation does not lead to a simple temperature dependence of average energy. This is because a) exponential term is both degree and temperature dependent, and b) the integral is degree dependent. However, we can analyse and simplify the low and high temperature limits.

High-temperature Limits At high temperature, $\beta \to 0$ and hence $e^x \to 1+x$. Hence, the average energy $\bar{U}$ behaves as

$$\bar{U} \to \frac{|V|^2}{\pi\varepsilon^2\beta^3}\int_0^{\varepsilon k} x dx = \frac{|V|^2}{2\pi}\cdot\frac{k^2}{\beta} \tag{25}$$

The corresponding node probability in Eq.(19) is

$$P(\beta,k) = \frac{|V|^2}{2\pi|E|}\cdot\frac{k}{\beta} \sim k\beta^{-1} \tag{26}$$

Low-temperature Limits At low temperature, $\beta \to \infty$ and hence $e^x \gg 1$. The average energy is given by

$$\bar{U} \to \frac{|V|^2}{\pi\varepsilon^2\beta^3}\int_0^{\infty}\frac{x^2}{e^x}dx = \frac{|V|^2}{\pi\varepsilon^2\beta^3}I_B(2) \tag{27}$$

where $I_B(2) = \zeta(3)\Gamma(3)$ is the Bose integral, where $\zeta(3)$ is a Riemann zeta function and $\Gamma(3)$ a gamma function.

Then, the corresponding node probability in Eq.(19) is

$$P(\beta,k) = \frac{C}{\varepsilon^2\beta^3}\cdot\frac{1}{k_T^2} \sim k_T^{-2}\beta^{-3} \tag{28}$$

where $C = |V|^2 I_B(2)/\pi$ is a constant, and $k_T = \omega_T/\varepsilon$.

5 Experiments and Evaluations

5.1 Data Set

Data Set 1: Here we use real world complex networks from the KONECT database. This database contains a variety of networks including

- The collaboration graph for authors of scientific papers from the arXiv's High Energy Physics-Theory (hep-th) section. Here an edge between two authors represents a common publication [9]. There are 22,908 vertices and 2,763,133 edges in the network.
- Facebook friendships network is the undirected network containing friendship of users. A node represents a user and an edge represents a friendship between two users [10]. There are 63,731 vertices and 817,035 edges.
- The Orkut network is the social network of Orkut users and their connections. There are 3,072,441 vertices and 117,185,083 edges. [11].
- The PPIs dataset extracted from STRING consisting of networks which describe the interaction relationships between histidine kinase and other proteins [13]. There are 216 vertices and 5,389 edges in the network.

Data Set 2: This data comes from the New York Stock Exchange. It consists of the daily closing prices of 3,799 stocks traded continuously on the New York Stock Exchange over 2619 trading days. The stock prices were obtained from the Yahoo! financial database [12]. A total of 415 stock are selected with the historical stock prices from the beginning of January 2010 to the end of June 2020. In the network representation, the nodes correspond to stock and the edges indicate that there is a statistical similarity between the time series associated with the stock closing prices.

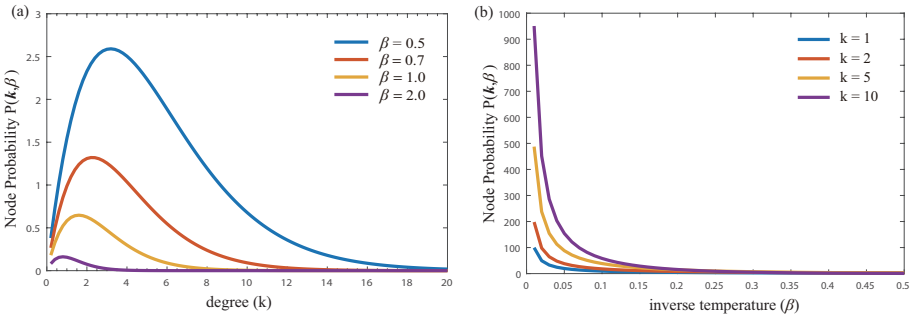

Fig. 1. The node probability varying with the degree k and inverse temperature β in Eq.(19). (a) node probability with degree; (b) node probability with inverse temperature

5.2 Experimental Results

We first conduct a numerical analysis on the node probability in Eq.(19). Fig.1 plots how the node probability varies with the degree k and inverse temperature

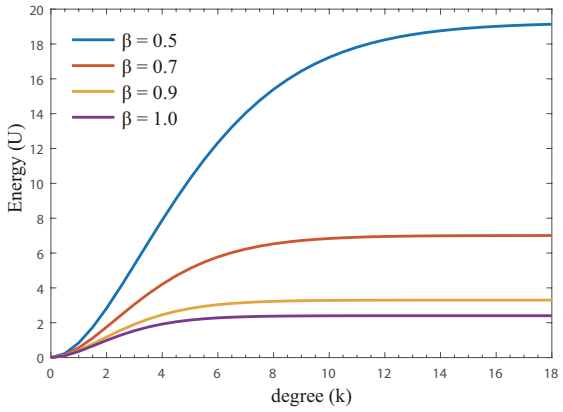

Fig. 2. Network energy varying with degree according to Eq.(18)

β, respectively. In Fig.1(a), there is a phase transition for the probability varying with the node degree. When the value of inverse temperature β increases, the peak corresponding to the phase transition shifts towards zero. In Fig.1(b), the node probability exponentially decays with the inverse temperature. The larger value of node degree, the faster in decay.

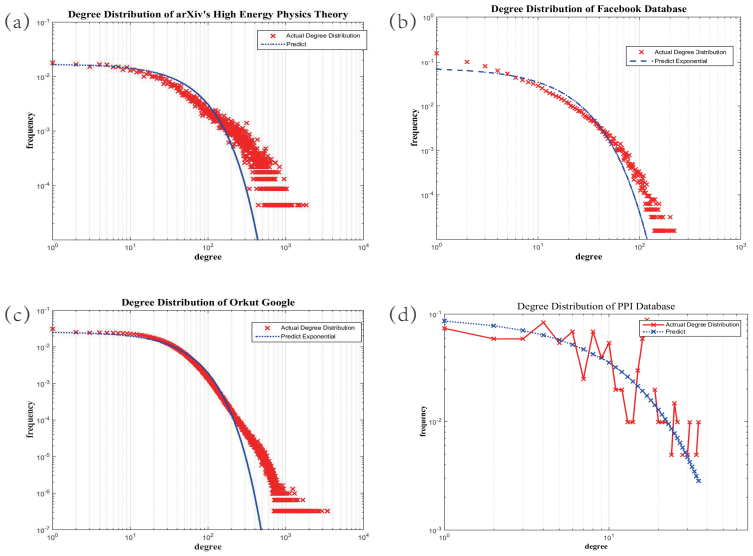

Fig. 3. Degree distributions for real-world networks. The red curves are the actual degree distributions and the blue curves are the result of simulation using Eq.(11).

Next we analyse the behaviour of the energy given in Eq.(18) with both degree and temperature. The expression in Eq.(18) is quite complicated and it is

not obvious by inspection how energy depends on temperature. This is because exponential term is both degree and temperature dependent and the integral is degree dependent. Fig.2 shows the full degree dependence for the energy. The energy increases with degree until reaching a plateaux value when the node degree is large. The energy also decreases rapidly when the inverse temperature β increases.

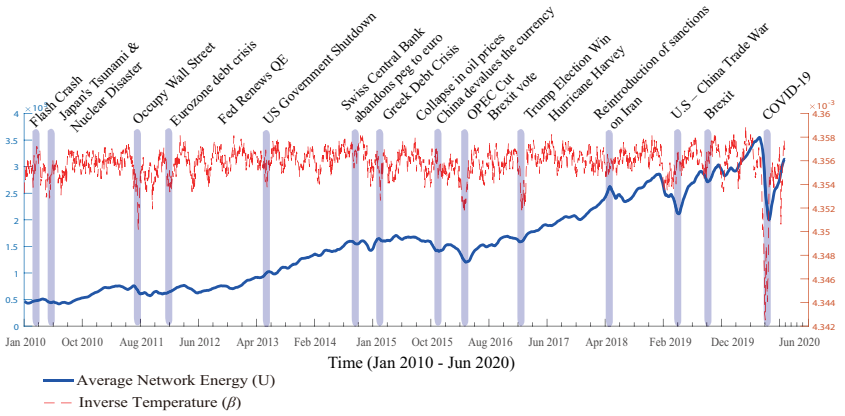

Fig. 4. The average energy and inverse temperature in S&P500 Index Stock Data (2010–2020) for original financial networks.

We now turn our attention to the real-world datasets. We examine the predictions of the node probability distribution in Eq.(11) for the complex network dataset. Fig.3 shows four degree distributions for different complex networks. The red curves are actual degree distributions and the blue curves are the predictions of our model. The four real networks come from the KONECT dataset, and are the arXiv hep-th network, the Facebook network, the Google Orkut user network, and a protein-protein interaction network. It is clear that, instead of following the power law degree distribution many real world complex networks follow the exponential distribution that we derived in Eq.(11). Actually, our model fits well at the low degree range, and at high-degree the power-law applies.

Finally, we use the time evolving financial networks to evaluate the energy and inverse temperature. Fig.4 plots the derived energy and inverse temperature for the stock exchange networks over the past decade. The stock market networks undergo rapid structural fluctuation during critical financial events. These events are listed in the caption of the figure. Compared with the energy, the temperature is more sensitive to the fluctuations of the financial markets. Sharp peaks in both energy and temperature indicate significant changes in network structure during the different financial events.

In summary, our derived expression for the degree distribution can therefore be used to fit the degree distributions of real complex networks. The corresponding energy and temperature associated with the network structure can also be used to identify abrupt changes of the pattern of edge connectivity.

6 Conclusion

In this paper, we make use of the Debye model to describe the node probability distribution in weighted networks. We commence from a thermal analogy described by the classical Boltzmann distribution. The particles in this thermal system are analogous to the nodes in a network. The energy is determined by the edge weights and node degree, which provides a physical interpretation for temperature. Then, the Debye solid model leads to an exponential expression for the probability density function of node degree. This the depends on the edge weights and the global temperature parameter, both related to the configuration of nodes and edges. The node probability density function together with the cumulative distribution function for energy reveal a phase transition for both the degree and temperature dependence. Experimental results show that the derived distribution can be used to fit the degree distribution in naturally occurring networks and identify the anomalous structure in time evolving networks.

References

1. Petri, G., Scolamiero ,M., Donato , I., and Vaccarino, F.: Topological strata of weighted complex networks. PloS one, vol. 8, no. 6, 2013.
2. Anand K , Krioukov D , Bianconi G . Entropy distribution and condensation in random networks with a given degree distribution. Phys Rev E, 2014, 89(6):062807.
3. Wang J, Lin C, Wang Y. Thermodynamic Entropy in Quantum Statistics for Stock Market Networks. Complexity. 2019 Jan 1;2019.
4. Cimini, G. , Squartini, T. , Saracco, F. , Garlaschelli, D. , Gabrielli, A. , Caldarelli, G.: The statistical physics of real-world networks. Nature Reviews Physics, vol. 1, no. 1, pp. 58–71, 2019.
5. Wang J, Richard W, and Edwin H, 2017. Spin statistics, partition functions and network entropy. Journal of Complex Networks, 5(6), 858-883.
6. Ye,C., Comin,C. H., Peron,T. K. D., Silva,F. N., Rodrigues,F. A., Costa,L. d. F., Torsello,A., and Hancock,E. R. Thermodynamic characterization of networks using graph polynomials. Phys Rev E, vol. 92, no. 3, p. 032810, 2015.
7. Wang J, Wilson RC, Hancock ER. Directed and undirected network evolution from Euler–Lagrange dynamics. Pattern Recognition Letters. 2020 Jun 1;134:135-44.
8. Jean-Charles,Delvenne,Anne-Sophie,Libert.Centrality measures and thermodynamic formalism for complex networks.Phys Rev E,2011,83:046117.
9. J. Leskovec, J. Kleinberg and C. Faloutsos. Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. ACM SIGKDD, 2005.
10. B. Viswanath, A. Mislove, M. Cha, Gummadi, Krishna P. On the Evolution of User Interaction in Facebook. Proc. Workshop on Online Social Networks. 2009, 37–42.
11. Yang J, Leskovec J. Defining and evaluating network communities based on ground-truth. Knowledge and Information Systems. 2015 Jan 1;42(1):181-213.
12. Silva FN, Comin CH, Peron TK, Rodrigues FA, Ye C, Wilson RC, Hancock E, Costa LD. Modular dynamics of financial market networks. arXiv preprint arXiv:1501.05040. 2015 Jan 21.
13. Szklarczyk D, Gable AL, Lyon D, et al, STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research. 2019 Jan 8;47(D1):D607-13.